



特許情報に関する言語生成モデルを 活用した知的財産創造手法の開発

Development of Intellectual Property Creation Method
Using Language Generation Model on Patent Information

Shigeaki Onoda

Graduate School of Information Engineering, Toyama Prefectural University
t855005@st.pu-toyama.ac.jp

Wednesday., 10 4, 2019,
Toyama Prefectural Univ.



前回まで

通常テキストマイニングではテキストを数値に表す方法として出現頻度のベクトルを使うがこれだと語彙数が増えるに連れて計算量も増えてしまうそこで計算量を節約するために固定長の入力ベクトルを使うことにした.

問題

学習済みの固定長単語ベクトルを使うときのベクトルサイズをあわせることが難しくモデルの作成で上手く組み込めなかったそのため詰まっていた.

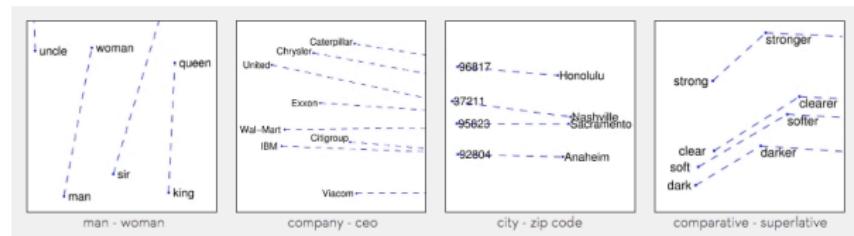
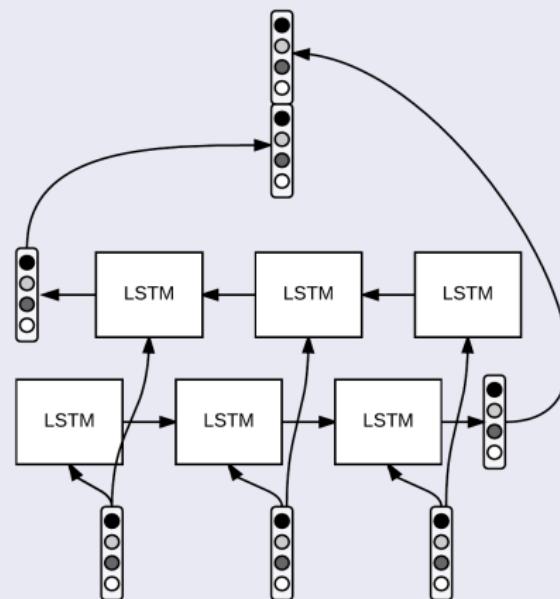


Figure: 1:glove モデルのイメージ



LSTM の双方化

精度を上げるために LSTM を前方向と後方向の Bi-directional 化を行った。





プロトタイプ

翻訳モデルをベースとしてモデルのプロトタイプを完成させた構成は以下図2

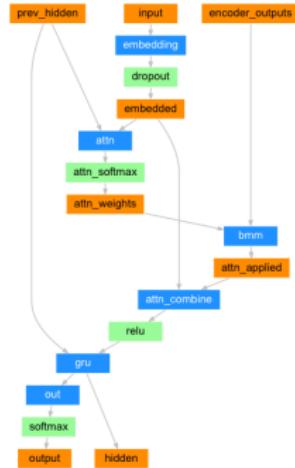


Figure: 2: プロトタイプの構成



構成

細かな設定項目

- 使用ネットワーク : LSTM, 注意機構
- 使用 word embedding : glove.840B.300d.model
web 記事から集めた 220 万語で 300 次元で構成される - 隠れユニット数 : 50
- 入力形式 : キーワード
- 出力形式 : 文章 (テストのため 10 文字以内)



現在の進度

小野田の想定

現在は超少量データでお試しの学習しているのみで本学習はしていない図3が現在のモデルから吐き出された混同行列(Confusion matrix)である実データで学習してないため精度は良くないが一応モデルが動くことが確認できた

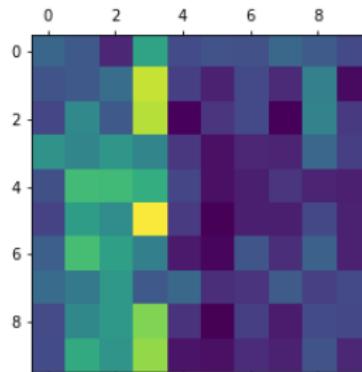


Figure: 3:結果



論文からのインスピアイア

前回発表した論文の使える部分

重要キーワードと記事からの要約文生成モデル

以下の論文ではキーワードが外部情報だが私の研究では特許のパラメータを扱う

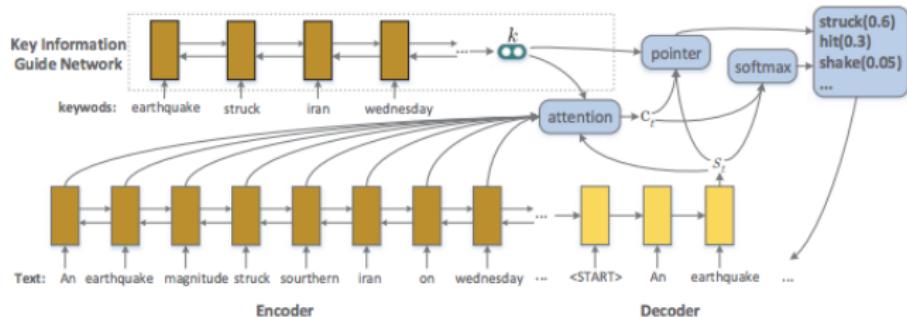


Figure: 4: ネットワークの概念図

URL: <https://www.aclweb.org/anthology/N18-2009>



今後やること

ToDo

- ① 学習データの充実
- ② 学習モデルの初学習
- ③ モデルのさらなる改良